Considering some of the practical issues encountered when finding and mapping wireless network access points.

802.11b Access Point Mapping

By Simon Byers and Dave Kormann

Scanning the electromagnetic spectrum for interesting non-natural signals has long been a pastime of curious hobbyists and professionals. In generic radio frequency signal discovery all of spectrum analysis, tuning, modulation, capture, and interpretation must be done to yield meaningful results. In the case of 802.11b all of this is done on the card itself, the results of which are made available by a supplied software interface—the card driver. This situation has opened up the pastime of radio discovery, interception and usage to unprecedented numbers of people, many of whom are not well-versed in radio frequency engineering or computer security. Other electromagnetic spectrum targets, such as mobile phones, satellites, and wireless video surveillance systems are more difficult to

AN ACTIVE CARD IN A REGULAR LAPTOP NOT SPECIFICALLY CONFIGURED FOR SCANNING IS PERFORMING THE SIGNAL RECOGNITION AND MEASUREMENT ALL THE TIME, THAT IS, IT IS COLLECTING DATA ON SIGNAL STRENGTH AND OTHER AP ATTRIBUTES THAT IT DISCARDS UNLESS TOLD OTHERWISE.

interact with and generally have been the domain of technically more proficient groups such as hackers, ham radio enthusiasts, and government organizations.

The fact it is so easy means the network discovery step itself is no longer particularly special. In fact it is hard *not* to find and connect to a wireless network in some places. For example, a laptop with a standard installation of Windows 2000 and a wireless card will aggressively connect to any open access point (AP) in its vicinity and, if possible, establish a connection to the Internet. This style of scanning provides an immediate benefit to the individual user, a benefit that may be extended beyond the immediate user by sharing information with associates and small groups. For us, though, the real value of AP finding and mapping comes from rigor in good data collection, meaningful visualization, and powerful analyses using any other ancillary data.

Related Work

There is a large body of work related to radio-frequency propagation and engineering; see, for example, [2, 6] or the issues of *IEEE Transactions of Antennas and Propagation* for technical discussion of radio frequency propagation. Goldhirsh [3] examines propagation and attenuation for mobile satellite receivers, which is an especially relevant topic when considering the mobile 802.11 receivers used in this work; propagation issues specific to 802.11a and 802.11b are discussed in [5] and [7].

Much software exists for mapping coverage and signal strength, including high-end tools such as those produced by EDX (www.edx.com) and free GPS interfaces like GPSDrive (www.kraftvoll.at/software/index.shtml). Projects such as WiFiMaps (www.wifimaps.com) are collecting the maps produced by these systems to provide a broader database of results. Combining land use data with digital elevation models in wireless coverage analysis is described in [4]; a higher-layer mapping effort in the spirit of this work is presented in [1].

Why Mapping?

Knowing the perimeter of one's network is fundamental from a general security standpoint. At the logical level, the perimeter of a network is defined by the physical extent of the network and its access controls. The introduction of wireless components to a network complicates this substantially due to the problems of establishing physical extent and the transitory and ill-defined nature of presence on a network in this domain.

General principles aside, the specific reason one is engaged in AP mapping affects the equipment required and the strategy employed. Among the motivations to map are:

- To find networks for the purposes of legitimate public use. An ordinary computer user seeking public Internet access might wish to scan for and select an acceptable nearby AP.
- To find networks to crack and abuse. The incentives to do this are numerous and might include, for example, four-letter content industry lobby groups attempting to break into the home machines of people they suspect are copyright infringers. More mundane motivations include theft of bandwidth or targeted network penetrations.
- To assess coverage of a particular network. For example, a network operator may wish to record, check, and publish the reach of a network.
- To assess coverage of networks in general. A network access company may wish to examine how many APs are present and what their characteristics are before deployment of a company's own APs.
- To find the existence of any networks in a particular small area. This may be required during an audit for rogue base stations inside a protected perimeter; such checks are carried out by companies with sensitive data that have policies regarding connection to their protected networks.
- To conduct a taxonomic analysis of network properties, specifically how many open networks

are there versus closed, how many are connected back to the Internet or which vendor is selling the most APs.
- To assess saturation of the spectrum. This is already an issue in certain areas and may become increasingly so as equipment becomes cheaper.

Each of these goals suggests different techniques: the abusive user will of necessity seek lower-profile, more inconspicuous mapping techniques than the authorized engineer. Larger surveys demand more organization while targeted measurements can be more comprehensive.

Software

Since Peter Shipley's original Perl script for BSD (www.dis.org/wl/wi-scan) many scanning and monitoring tools have been published, and some are sold. Two big names are NetStumbler (netstumbler.com) and Kismet (kismetwireless.net), which run on Windows and Linux/BSD, respectively. We will avoid debate on the packages themselves due to fast development in the field and will instead concentrate on requirements we consider important.

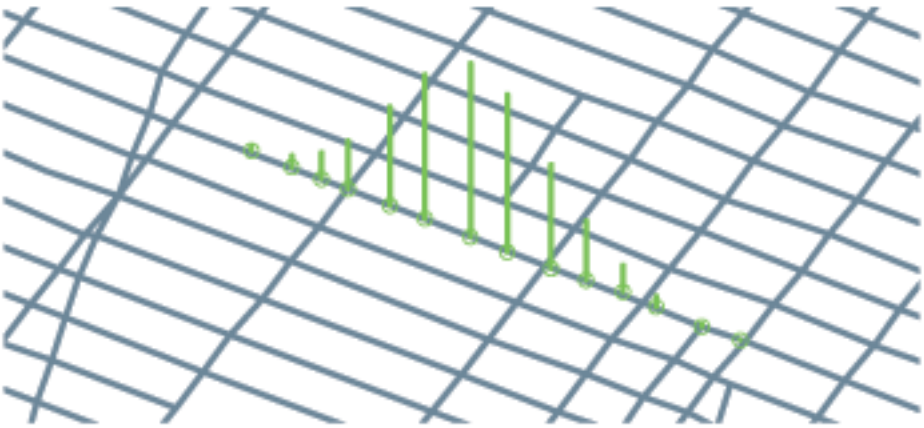

Figure 1. An informative network encounter. The heights of the bars show the signal-noise ratio. Note the samples show a recognizable decaying coverage area.

The basic items to log are AP name, MAC address, signal, noise (and hence SNR), channel, and encryption. The correlation of this data with spatial coordinates is what allows mapping, although we recommend logging of sampled locations even when there is no signal present.

When assessing networks it is possible to conduct more intrusive detailed tests that verify openness of the network by demonstration. A traceroute to a known point would both validate connectivity and yield information about the nature of the connection, for example, whether it is DSL or cable broadband.

Conversely some useful information can be gathered using entirely passive means. Analysis of traffic obtained by standard sniffing tools can help to locate APs that are not broadcasting their SSID, and can assist in determining whether an AP is a useful path to the Internet or only to a LAN.

We also note the possibility of software countermeasures in this area. For example, counter scanning techniques have recently been developed involving generation of false AP strikes when active scanning is detected. Such effects may also occur inadvertently, as when two active scanners erroneously detect each other as APs.

Legalities and Ethics

When scanning we have always carefully avoided making IP connections on networks we do not have permission to access. While we have not received expert advice on the legality of simply mapping networks, it seems possible to draw a clear distinction between noting signal strength in the public airspace and exploiting those airwaves to access a network without consent. Further, an active card in a regular laptop not specifically configured for scanning is performing the signal recognition and measurement all the time, that is, it is collecting data on signal strength and other AP attributes that it discards unless told otherwise. This is without taking into account the connection that will occur if a discovered AP is open and offering DHCP.

Traffic analysis, of course, raises its own set of issues, even on unencrypted links. Passive interception of data transmission may expose a listener to the measures of the Electronic Communications Privacy Act. At the very least, it seems that anything beyond routine signal-strength analysis should be done with caution and under controlled conditions. Again, unintentional trespassing might be difficult to avoid: Kismet, in particular, functions by intercepting traffic in order to discover topology. This fact alone renders its use on public networks dangerous.

Hardware

Some form of computer device is required to control and listen to the wireless card. Custom

802.11b scanning handhelds exist but they are not very flexible for advanced data analysis and are somewhat expensive. Either a laptop or a handheld computer is normally used, laptops being more associated with work done in vehicles. Handhelds, however, have an advantage for surreptitious or indoor scanning. Generic PDAs function adequately in this role and are sufficiently unobtrusive to not draw attention to themselves, but the data they collect must be transported to a more capable machine for advanced analysis.

Antennae

A substantial core of everyday 802.11b users carry an antenna for better radio connections in normal usage. Often these are small omnidirectional antennas but directional versions are also used. Even if physically quite large, if an antenna makes the difference between broadband connectivity and none at all, then for some users that justifies the extra bulk and complication.

In day-to-day use of wireless cards external antennae are the exception rather than the rule, but for dedicated scanning activities the reverse is true. The type of scanning affects the choice of antenna. When there is a specific target network or point location it is beneficial to use a directional antenna. A Yagi type is somewhat discreet (which presumably is a feature for those involved in the less-legal activities), while a dish will have a tighter beam and hence better connection but be unwieldy and obvious.

In more general network finding and automated data gathering an omnidirectional antenna is usually used, unless the work is done to incorporate knowledge of the scan direction at any given time into the data-gathering software.

Combinations of antennas may prove useful; for example in urban canyons an upward-scanning antenna in conjunction with an omnidirectional scan at ground level may prove effective. This is useful because a high-gain omnidirectional antenna has a coverage region that spreads out in two dimensions like a thin disk with very little visibility upward anywhere, least of all at its own location.

Location Recording Issues

A GPS unit is the best location-measuring device in outdoor environments. In highly urban areas satellite reception can be a serious problem so we usually use a high-gain external antenna. Hence the GPS should have both data output facility (a serial port connection) and an external antenna capability. Another optional but desirable feature is the ability to power the GPS externally.

Indoor location finding is much more difficult; we describe two methods we have implemented here. We have used interpolation in conjunction with fixed reference points by taking GPS readings at the corners of a building and used these coordinates to anchor the endpoints of walks down corridors. Using an alternate method, the spatial reference points can be specified by a point-and-click method, ideally using the CAD diagrams for the building in question as a base map. The radio frequency measurements are correlated with spatial locations by the use of timestamps output by the scanning software and the mouse clicks.

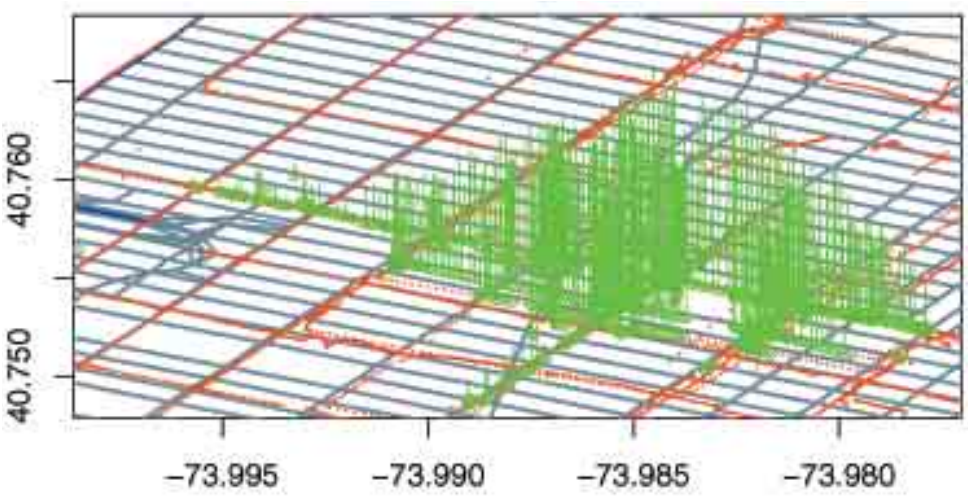

Figure 2. Map showing SNR of the NYCWireless node in Bryant Park, Manhattan.

Some Practicalities

In a detailed survey of a particular site, for example a corporate network, it is advisable to have the permission of the building manager before driving through the flower beds. A space-filling sample path should be (approximately) constructed to cover all areas of anticipated reception and those just outside or on long lines of sight.

In any type of wide area general survey a sampling scheme must be used. It is not possible to drive every street in Manhattan, for example, without spending a considerable amount of time and money. We try to use a scheme that covers quickly on the large scale but allows calibration by detailed sub-samples. Repeated measure samples are often required and should be cal-

ibrated for the day of week and the time of day due to the switching off of some networks at night and weekends. Planning involving a map and knowledge of one-way streets can yield quite rapid and comprehensive coverage sampling schemes by the generation of boustrophedonic drive paths and/or minimization of left turns (in countries that favor the left-hand drive protocol). Traffic is a significant impediment to sampling in a vehicle, hence our preference for performing driving work between midnight and 6 A.M., but we note that some sampling should be done during traditional business hours.

When sampling in a vehicle an AC-to-DC power inverter is useful for powering the laptops and optionally the GPS and other peripherals. We note also that in hot climates air conditioning may be required to cool the laptops.

Our aim when driving is to drive through the average network at such a speed that we get several samples from it, as demonstrated in Figure 1. These samples have enough structure to allow some meaningful reconstruction of the coverage field in post processing. Having only two or three sample points gives highly variable estimates of the location of the physical AP and its coverage. Moving at 20 to 30 miles per hour seems to do a good job of this in many situations, but this can be adapted to local conditions. Higher speeds can be used well on open roads with very little structure around: for example even at 65 MPH one can collect many sample points from certain military bases in the desert a mile or so from the roadway.

Results

General Statements. The general result of scanning for APs is that wherever you look you will find something. Dense urban areas offer thousands of APs, suburbs hundreds, and even a drive on the highway through seemingly empty countryside results in hits from towns at a distance of a mile or more. We have witnessed explosive growth in APs in all of these settings in the past 18 months.

AP Uses. The usage of AP equipment varies from place to place and between businesses and residences. In mixed areas such as Manhattan this distinction is not as obvious but in suburban areas there is a sharp change in the nature of 802.11b as you turn away from a strip mall and move into quiet streets filled with housing. Businesses tend to use a name for their AP that matches their own name and also tend to use WEP. They also are likely to use the higher powered equipment such as the Cisco 100mW base stations. In residential neighborhoods it is not uncommon to find a predominance of less powerful (30mW) base stations, for example Linksys, with default settings in place (no access control or encryption) named Linksys and offering DHCP.

Manual examination of this data can reveal some of the uses to which 802.11b equipment is being applied. These diverse tasks include truck weigh stations on California freeways, rental car agencies in general, police stations in New Jersey, certain U.S. Marine Corps bases, and many hospitals. A drive targeting the premises of 802.11b equipment manufacturers can yield data worth examining in some detail.

Coverage of Free Public Access Point. The nonprofit group NYCWireless operates a free public node in Bryant Park, Manhattan. This node has three Cisco APs linked to the Internet via a T1 connection. This makes the park more attractive for business users and casual surfers but it is also interesting to know exactly how far the AP reaches beyond the park. Figure 2 shows a map with signal-noise ratios plotted at sampled points.

Figure 3. Map showing a selection of coverage given by Linksys APs in Manhattan left in default configuration.

It is obvious from the map in Figure 2 that the node is covering much more than just the park. We estimate that with a small Yagi antenna the node could be used from the front of buildings two or more blocks away in some directions, particularly to the west. We also note that on the east end of the park is the New York Public Library, a large stone building, behind which there is still significant signal. This hints at the complexities and usefulness of signal bounce in urban environments. Information such as this can feed directly back into the design and maintenance of such enterprises.

Further Analyses

Breakdown of AP Attributes. Of interest is the breakdown of APs by configuration. A large percent-

age of APs is invariably found to have the vendor-supplied default configuration, that is, no WEP, default AP name, and SSID broadcast enabled. Usually only a quarter of APs have WEP enabled and up to half in some areas have default SSIDs. The network of APs formed by those from a particular vendor and left in default configuration is of particular interest as adjacent APs with identical configurations greatly facilitate a type of ad hoc roaming. Figure 3 shows such an accidental network formed by APs from one particular vendor.

Locations of APs. In theory, given spatially separated measurements of signal strength, an estimate of the location of an AP can be formed using the intersecting spheres problem. However, in practice this is not useful as many drive paths are approximately linear and many network strikes fall only on one of these linear segments. Many people (including ourselves) use the point at which maximum signal-noise ratio is achieved as a simple-to-compute estimate of AP location.

Using external data such as names and addresses from the telephone book can prove quite useful in locating an AP. First, we generate spatial coordinates for all the addresses in the area of interest (a city, for example) by using the U.S. Census Tiger data. Then, for a given network strike we trawl through the spatially closest name-address pairs using approximate string matching to associate the AP with a physical address and hence a point in space (possibly in three dimensions). This works when people name their AP after their business, themselves, or even after their physical address, but the proportion of APs for which this is useful is low, less than 10% even in Manhattan where it works best. Despite this, correlating an AP with its owner is a desirable goal in itself.

Interaction with External Data. Using the type of name-address data from the previous section, it is possible to construct population coverage analyses given the AP data. For example, for each of the approximately 800,000 addresses appearing in the Manhattan phone book it is a simple matter to assess whether a particular address seems to be within range of a known open residential AP with default settings. Hence we determine how many people could potentially surf for free on existing infrastructure by using, say, a directional antenna and a wireless card.

This technique when applied to commercial 802.11b networks is even more interesting as it yields not only estimates of potential customer base for a given deployment but also the list of prospects. Further, if we estimate the telephone wire distance of each address, fold in the DSL availability as a function of telephone area code and prefix, and compare that with the wireless availability estimate we can begin to explore how different last-mile alternatives might compete or complement each other on the scale of cities. There is some evidence these techniques are already being used by residential broadband companies to determine whether subscribers are in violation of their service agreements by offering public access via their broadband connection.

On a less commercial note, the data we collect is used to assess the density and availability of the edge of the Internet. In general, network access can be via dialup, DSL, cable, T1, or even other wireless devices, and incorporating the 802.11b coverage into the global picture of network access is an interesting area of study that we have some experience with. In practical terms each point of access is a potential attack launch point, but conversely and especially in the case of wireless, a potential local re-routing resource in the event of disruption closer to the center of the network, such as happened in September 2001.

Conclusion

Due to the proliferation of wireless cards and mobile computing devices, AP mapping is both easy to do and useful once it is performed. The full range of informal network finding from a scan from a single apartment to detailed mapping of a city can be done with relatively few resources and yields tangible short- and long-term benefits to those who can use the data thus generated. When wireless coverage data is collected with a certain degree of rigor we have found that sufficient quality and scale can be achieved to permit use in cutting-edge data mining activities that yield meaningful economic, social, and technological results. C

References

1. Cheswick, B., Burch, H., and Branigan, S. Mapping and visualizing the Internet. In *Proceedings of the USENIX Technical Conference*, 2000.
2. Griffiths, J. *Radio Wave Propagation and Antennas, An Introduction.* Prentice-Hall, 1987.
3. Goldhirsh, J. and Vogel, W.J. *Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems.* NASA Reference Publication 1274, Second Edition.
4. Kirner, J.L. and Anderson, H.R. The application of land use cover data to wireless communication system design. In *Proceedings of the ESRI User Conference*, 1998.
5. Nabritt, S.M. Modeling multpath in 802.11 systems; www.eetimes.com/story/OEG20021008S0001.
6. Rappaport, T.S. *Wireless Communication, Principles and Practice*. Prentice-Hall, 1996.
7. Yee, J. and Pezeshki-Esfahani, H. Understanding wireless LAN performance trade-offs. *Communication Systems Design* (Nov. 2002).

Simon Byers (byers@research.att.com) is a senior member of the research staff at AT&T Laboratories in Florham Park, NJ.
Dave Kormann (davek@research.att.com) is a senior member of the research staff at AT&T Laboratories in Florham Park, NJ.